A COMPARATIVE APPROACH TO SPECIES DIVERSITY

by

Daniel L. Solomon

Abstract

A partial order (majorization) is defined on the set of species abundance vectors corresponding to a collection of communities which is recommended for ordering those communities according to diversity. The use of a partial order admits the possibility that a pair of communities may have different species abundance patterns but not be comparable with respect to diversity. It is also noted that the commonly applied diversity indices preserve this order so that a contradictory ordering of two communities by two such indices can be interpreted as non-comparability of the communities under the partial order. Finally, the partial order naturally suggests the "expected abundance rank" of an individual selected at random from the community as a diversity index.

INTRODUCTION

We are interested in characterizing biological (or sociological) populations which we shall call communities. The elements of a community are assumed classifiable by some characteristics into subpopulations to which we apply the term species. The study of the diversity of a community is the study of the number of its species (the community's species-richness) and their relative abundance (called variously evenness, equitability or dominance). Diversity increases with richness

---

Paper No. BU-573-M in the Biometrics Unit Mimeo Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853.

and evenness. The term species is to be interpreted broadly, including the usual taxonomic definition but also classifications based on other criteria, e.g., "tall", "medium", and "short" as tree "species", or "butcher", "baker", and "other" as human "species". Similarly, "abundance" includes, for example, numbers of individuals for animal species and biomass for phytoplankton.

The species abundance characteristics of an s-species community can be described by a vector $\underset{\sim}{p} = (p_1, p_2, \cdots, p_s)$ where p_i is the relative abundance of the i^{th} species, e.g., the proportion of individuals in a community who are members of the i^{th} species. Thus $\underset{\sim}{p}$ is a probability vector. For purposes of this discussion, we ignore the difficult sampling problems associated with studies of diversity, assuming that the species abundance vector from a census of the entire community is available. Ecologists and others have adopted a number of indices of the diversity of such a community, the most commonly applied of which are simple functions of (s), the community's species-richness, $S(\underset{\sim}{p}) = 1 - \Sigma p_i^2$ which is a variation on Simpson's Index, and $H'(\underset{\sim}{p}) = -\Sigma p_i \ln p_i$ called the Shannon-Weaver or Information Index. Note that the latter indices (those of the evenness component) are maximized for given s at $\underset{\sim}{p} = (1/s, 1/s, \cdots, 1/s)$ and minimized at any permutation of $(1, 0, 0, \cdots, 0)$.

To this worker's knowledge, nowhere in the biological literature has a fundamental <u>definition</u> of diversity been given – the concept has only been "defined" by its measures, the diversity indices. This is akin to "defining" I.Q. as that quantity measured by I.Q. tests. An implication of "defining" diversity by indices is that given a numerical valued (quantitative) index, every pair of communities can be compared with respect to diversity. This is not a biological imperative. That is, presented with differing species abundance lists for a pair of communities, we may not wish to identify one as more diverse. The objective of the work described here is to make a contribution to the establishment of a notion of diversity, more fundamental than that associated with diversity indices, investigating measures which are comparative or relative rather than quantitative.

THE DOMINANCE COMPONENT

We begin by (temporarily) restricting attention to the dominance component of diversity, assuming that we are interested only in communities with some fixed number, s, of species. We now define a partial order on the set of (ordered) probability vectors of dimension s. Thus let

$$\mathcal{P} = \{\underset{\sim}{p} \in \underset{\sim}{\mathbb{R}}^s | 1 \geq p_1 \geq \cdots \geq p_s \geq 0;\ \Sigma p_i = 1\}$$

represent the set of species abundance vectors for an s-species community, where p_1 is the relative abundance of the most abundant species. Note that in studying species diversity (as opposed to community similarity), the labels associated with the species are irrelevant, e.g., a community with 3 representatives of species A and 2 of species B is as diverse as one with 2 of species A and 3 of B. More precisely, we assume that there is an importance value scale of equilibrating units, e.g., biomass.

We shall write for $\underset{\sim}{p}$, $\underset{\sim}{q} \in \mathcal{P}$,

$$\underset{\sim}{p} \succ \underset{\sim}{q} \quad \text{read "}\underset{\sim}{p}\text{ majorizes }\underset{\sim}{q}\text{"}$$

to mean $\sum_{i=1}^{j} p_i \geq \sum_{i=1}^{j} q_i$, for each $j = 1, 2, \cdots, s - 1$. For example, $(2/3, 1/4, 1/12) \succ (1/2, 1/3, 1/6)$. It is not difficult to show that for every $\underset{\sim}{p} \in \mathcal{P}$, we have

$$(1, 0, 0, \cdots, 0) \succ \underset{\sim}{p} \succ (1/s, 1/s, \cdots, 1/s).$$

Thus majorization orders s-tuples according to dominance (backwards according to evenness), and in what follows, we will try to motivate the use of majorization to order communities.

It can be shown that majorization is a partial order so that if $\underset{\sim}{p} \succ \underset{\sim}{q}$ and $\underset{\sim}{q} \succ \underset{\sim}{r}$ then $\underset{\sim}{p} \succ \underset{\sim}{r}$. However, it is not true that for every $\underset{\sim}{p}, \underset{\sim}{q} \in \mathcal{P}$, either $\underset{\sim}{p} \succ \underset{\sim}{q}$ or $\underset{\sim}{q} \succ \underset{\sim}{p}$. For example, $\underset{\sim}{p} = (.4, .2, .2, .2)$ and $\underset{\sim}{q} = (.33, .33, .20, .14)$ are not comparable. If these represent relative abundance patterns for two 4-species communities, which "should" be termed less diverse (more dominated)? To answer this question, we might calculate some common diversity indices for the communities, but in so doing would find: $H'(\underset{\sim}{p}) > H'(\underset{\sim}{q})$ but $S(\underset{\sim}{p}) < S(\underset{\sim}{q})$. We might conclude that the first community is strongly dominated by one species but is otherwise even, while the second community is moderately dominated by two species. We thus would not like to insist that either community is more diverse. As we shall shortly see, the fact that H' and S order the two communities differently is a reflection of the non-comparability under majorization of $\underset{\sim}{p}$ and $\underset{\sim}{q}$. Note too that even when two indices order a pair of communities in the same way, they will in general disagree about the extent to which one community is more diverse than the other. Thus, it is not clear which, if any, of the common indices is in a meaningful scale for comparing communities.

We shall next demonstrate that the proposers of diversity indices have been concerned (no doubt unknowingly) about majorization. To do so, we begin by making the following definition[1]:

With $\mathcal{P}$ as previously defined, a real valued function $f:\mathcal{P} \to \underset{\sim}{\mathbb{R}}^1$ for which (the partial derivatives exist and)

$$\left(\frac{\partial f}{\partial p_i} - \frac{\partial f}{\partial p_j}\right)(p_i - p_j) \leq 0 \text{ for all } p_i, p_j$$

is called <u>S-concave</u>.

---

[1] A more general definition of S-concavity is available which does not require existence of the derivatives. We have used the more readily applied version as a working definition. See Berge [1963] for a more careful treatment.

For example, for Simpson's index, $S(\underset{\sim}{p}) = 1 - \Sigma p_i^2$, we have $\frac{\partial S}{\partial p_i} = -2p_i$ whence $\left(\frac{\partial S}{\partial p_i} - \frac{\partial S}{\partial p_j}\right)(p_i - p_j) = -2(p_i - p_j)^2 \leq 0$ for all p_i, p_j and so S is S-concave. An important subclass of the class of S-concave functions includes all functions of the form $f(p_1, \cdots, p_s) = \sum_{i=1}^{s} \phi(p_i)$ where ϕ is a concave (in the usual sense) function of a single variable.

It is not difficult to check that <u>all the common diversity indices (or their negatives) are S-concave</u>. This includes the indices due to Simpson, Shannon, McIntosh, Pielou, Lloyd and Ghelardi, and Brillouin. The significance of this observation rests in the following:

The function f is S-concave if and only if $f(\underset{\sim}{p}) \leq f(\underset{\sim}{q})$ whenever $\underset{\sim}{p} \succ \underset{\sim}{q}$.

Thus, the S-concave functions constitute precisely the class of functions which preserve order (backwards) under majorization.

An implication of the S-concavity of diversity indices is that if $\underset{\sim}{p} \succ \underset{\sim}{q}$, the indices will all order the corresponding pair of communities in the same way. Equivalently, it is only if $\underset{\sim}{p}$ and $\underset{\sim}{q}$ are not comparable that one can expect to find two diversity indices which order the communities differently.

Thus we conclude that majorization can be viewed as a comparative measure of dominance so that $\underset{\sim}{p} \succ \underset{\sim}{q}$ can be interpreted: "An s-species community with species abundance vector $\underset{\sim}{p}$ is more dominated than one with vector $\underset{\sim}{q}$."

THE EXPECTED ABUNDANCE RANK INDEX

Motivated by the observation that majorization is a common thread behind the usual diversity indices, perhaps we can establish a <u>relative</u> measure of dominance, tied to the "extent" by which one community is majorized by another. For example, if $\underset{\sim}{p} \succ \underset{\sim}{q}$, we might measure their "diversity distance" by (recall the definition of majorization)

$$\|\underset{\sim}{p} - \underset{\sim}{q}\| \equiv \sum_{j=1}^{s-1}\Big[\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} q_i\Big] = \cdots = \sum_{i=1}^{s} iq_i - \sum_{i=1}^{s} ip_i .$$

Thus this distance between two communities is determined by the difference between their Σip_i. This suggests some function of $R \equiv \Sigma ip_i$ as a numerical diversity index. Note that R is the <u>expected abundance rank</u> (recall $p_1 \geq p_2 \geq \cdots \geq p_s$) of an individual drawn at random from the community. For a highly dominated community this should be small (many rank 1 individuals) whereas for an even community, it will be large. Thus R is an evenness or equitability index. The range of the index for given s is $1 \leq \Sigma ip_i \leq (s + 1)/2$ and so (as is done for other indices) this index can be standardized for s by forming $I \equiv [\Sigma ip_i - 1]/[(s + 1)/2 - 1]$ so that $0 \leq I \leq 1$ for all s.

Note too that Σip_i is S-concave (as is I) but not of the form: $\Sigma\phi(p_i)$ for some concave function ϕ, as are the common diversity indices.

THE RICHNESS COMPONENT

To this point we have only considered pairs of communities with the same number of species. That requirement can be weakened as follows. Typically, we wish to compare communities that are reasonably similar; e.g., a specific community before and after some perturbation of the environment. Thus we might conceive of a maximum number, s, of species that <u>could</u> be present. Then all species abundance vectors would have s components but (perhaps) have zeros corresponding to missing species. Majorization still applies with the following implication: If

$$\underset{\sim}{p} = (p_1, p_2, \cdots, p_t, 0, \cdots, 0),\ p_t > 0,$$

and

$$\underset{\sim}{q} = (q_1, q_2, \cdots, q_u, 0, \cdots, 0),\ q_u > 0 \text{ and } s \geq t > u$$

so that the $\underset{\sim}{p}$ community has more species present, then $\sum_{i=1}^{u} p_i < 1 = \sum_{i=1}^{u} q_i$ and $\underset{\sim}{p} \not\succ \underset{\sim}{q}$.

Thus, in ordering the diversity (richness and evenness) of communities by majorization, for a given pair of communities, either the community with the larger number of species is more diverse or the communities are not comparable. (Recall that majorization orders communities backwards with respect to evenness.)

SUMMARY

In summary, we have presented a partial order on the set of species abundance vectors which we recommend for ordering communities according to their diversity. The use of a partial order admits the possibility that a pair of communities may have different species abundance vectors but not be comparable with respect to diversity. It has also been observed that the commonly applied diversity indices preserve this order so that the disagreement in ordering two communities, possible when computing two indices for each, can be interpreted as non-comparability of the communities. Finally, this partial order suggests the expected abundance rank of an individual selected at random from the community as a diversity index whose properties should be investigated.

Reference

Berge, C. [1963]. Topological Spaces. MacMillan, New York.